Converting text input into moving-face picture **Best Available Copy**

Patent Number: F GB2231246

Publication date: 1990-11-07

HIGUCHI NORIO:: KOIKE ATSUSHI;; KANEKO MASAHIDE;; HATORI YOSHINORI;; Inventor(s):

YAMAMOTO SEIICHI

KOKUSAI DENSHIN DENWA CO LTD (JP) Applicant(s):

Application

GB19900005142 19900307 Number: Priority Number(s): JP19890053899 19890308

IPC Classification: G06F15/72

EC Classification: G10L21/06L, G06T9/00F, G06T15/70, G10L9/20, H04N7/26J4

Equivalents: JP2518683B2

Abstract

A moving picture of a face with mouth-shape variations corresponding to a text sentence input is produced. The input sentence is divided into a train of phonemes and a speech synthesis technique capable of outputting a voice feature of each phoneme and its duration is utilized. Based on the voice feature, a mouth-shape feature corresponding to each phoneme is determined 3. Based on the mouth-shape feature, the value of a mouth-shape parameter is determined 5, 4 for representing a mouth shape. Further, the value of the mouthshape parameter for each frame of the moving picture is controlled 2 in accordance with the duration of each phoneme, thereby synthesizing the moving face picture having mouth-shape variations which agree with the

speech output.

Data supplied from the esp@cenet database - I2

① 特許出顧公開

⑩ 公 開 特 許 公 報 (A) 平2-234285

⑤Int. Cl. ' G 06 F 15/72 G 10 L 3/00 9/20 識別記号 350 庁内整理番号 8125-5B 個公開 平成2年(1990)9月17日

5 0 8125-5B S 8622-5D A 8622-5D

審査請求 未請求 請求項の数 3 (全9頁)

1

❷発明の名称 画像合成方法及びその装置

②特 額 平1-53899

②出 願 平1(1989)3月8日

@発 明 者 金 子 正 秀 東京都新宿区西新宿2丁目3番2号 国際電信電話株式会 社内

⑰発 明 者 小 池 淳 東京都新宿区西新宿2丁目3番2号 国際電信電話株式会

社内

⑰発 明 者 羽 鳥 好 律 東京都新宿区西新宿2丁目3番2号 国際電信電話株式会

社内

⑩発 明 者 山 本 誠 一 東京都新宿区西新宿2丁目3番2号 国際電信電話株式会

社内

⑪出 願 人 国際電信電話株式会社 東京都新宿区西新宿2丁目3番2号

⑩代 理 人 弁理士 大 塚 学 外1名

最終頁に続く

明 細 書

1. 発明の名称

画像合成方法及びその装置

- 2. 特許請求の範囲

(2) 文字列として表現される文章を入力するため の入力端子と、該入力端子より入力される該文 字列を音素列に分割し、各音素ごとに音声特徴 及び持続時間を出力することが可能な音声合成 部と、各音索ごとの該音声特徴から口形特徴へ の変換を行う変換部と、種々の口形特徴と具体 的な口形状を表現する口形状パラメークとを対 応付けた変換テーブルと、前記変換部で得られ る各音素ごとの口形特徴に対応する口形状パラ メータを前記変換テーブルから取出す口形状パ ラメータ取得部と、一定時間間隔の画像系列と して与えられる動画像を生成するために該口形 状パラメータ取得部から得られる口形状パラメ ータの値の出力を前記音声合成部から与えられ る各音器ごとの持続時間に従って側御するため の時間調整部と、該時間調整部の制御のもとに 前記口形状パラメータ取得部から出力される口 形状パラメータの値に従って画像を生成するた めの画像生成部とを備えたことを特徴とする画 像合成装置。

3. 発明の詳細な説明

(発明の技術分野)

本発明は、ディジタル処理による画像合成方法 に関するもので、特に、発声に伴う口形状変化を 表現する顔画像(静止画像または動画像)を合成 する方式に関するものである。

口形パターンの対応付け部、52は音節と口形パタ ーンの対応テーブル、53は口形状選択部、54は口 形状用メモリである。次に各部の動作を簡単に説 明する。音節分離部50は、入力された文章(文字 列)に対して、これを音節単位に区切る働きをす る。例えば「kuna」という入力は、[ku]と[ma]の 2 つの音節に分けられる。次に、音節と口形パタ ーンの対応テーブル52は、予め用意された音節と 口形パターンの対応関係を蓄積したテーブルであ る。 音郎は "a", "ka" などひとまとまりの音 を表現するものである。 口形パターンは、大口形 (<A><I><U><E><K>等)と小口形 (くu>くo>くk>くs>等) とがあり、口形 の種類を示すものである。これらを用いて"a" に対しては<A><*><A>, "ka" に対して は<K><*><A>というように音節と口形パ ターンの対応関係をテープルにしておくわけであ る。ごこで、<*>は中間口形を示す。音節と口 形パターンの対応付け部51では、音節分離部50か ら送られてくる一つ一つの音節ごとに、音節と口

(従来技術)

人が発声する場合には、調音器官により音声情報が生成され、同時に、外見的な変化として発力に伴い、口部分の動き(形状変化)が生じる入力をが直接発声するのではなく、文字列として入入して入た文章を音声情報に変換して出力する方法は音声合成と言われ、従来、多くの成果が得られてきている。一方、入力された文章に対して対応するして形状変化を有する顔画像を生成する方法に関しては従来技術は少なく、松岡清利、黒須顕二による次の報告があるにどどまっている。

松岡、黒須の方法は、(松岡淯利、黒須顕二:「聴覚障害者の説話訓練のための動画プログラム」電子情報通信学会論文誌、vol.J70-D.no.11.PP. 2167-2171(1987年11月))に示されている。これは、プログラムの形で実施されているが、入力された文章に対して、対応する口形状変化を得るための考え方の基本を整理して示すと、第6図のようになる。

第6図において、50は音節分離部、51は音節と

この外に、関連する従来技術として、文章入力に対してではないが、音声を入力として対応する 口形状変化を推定する方法も報告されている。これは、〔森島繁生、相沢滑晴、原島博:「音声情報 に基づく表情の自動合成の研究」第4回NICOGRAPII 論文コンテスト論文集、PP.139-146、日本コンピュータ・グラフィックス協会(1988年11月)〕に 示されている。ここでは、入力された音声情報に対して、対数平均パワーを計算して口の聞き具合を制御する方法と、声道のホルマント特徴に対応する線形予測係数を計算して口形状を推定する方法と2週りが提案されている。

(発明が解決しようとする課題)

本発明の第2の特徴は、文字列として表現される文章を入力するための入力端子と、 該入力端子より入力される該文字列を音素列に分割し、各音 常ごとに音声特徴及び持続時間を出力することが可能な音声合成部と、各音業ごとの該音声特徴から口形特徴への変換を行う変換部と、 種々の口形特徴と具体的な口形状を表現する口形状パラメー

固定的に4フレームを割当てており、入力される 文章に応じた自然な口形状変化を表現することが 困難であるという問題がある。また、入力された 文章に対して、音声と口形状画像を同一のタイミ ングで出力しようとした場合に両者でのマッチン グをとることも困難である。

さらに、森島、相沢、原島の方法は、入力された音声情報をもとにして口形状を推定するという技術であり、文章を入力して、これに対応した口形状変化を有する動画像を生成するという目的には適用することができない。

(発明の目的)

本発明は、上述した従来技術の問題点を解決するためになされたのもであり、音声出力との的確な対応付けがなされ、かつ、各音繁の持統時間に合わせた形で口形状変化を表現することが可能な画像合成方法及びその装置を提供することを目的とする。

(発明の構成)

本発明の第1の特徴は、文字列として表現され

タとを対応付けた変換テープルと、前記を投いた、前記を投いたとの口形特徴に対応を投いている各音器記を投テープルからの取出を表示の口形では、一夕取得部と、大きなのになって、一夕取得部では、一夕取得のでは、一夕取得では、一夕取得では、一夕取得では、一夕では、一夕取得では、一夕ない。

本発明の第3の特徴は、文字列として表現され、 る文章を入力するための入力端子と、該人力端子 より入力される該文字列を音緊列に分割し、各音 衆ごとに音声特徴及び持続時間を出力することが 可能な音声合成部と、各音素ごとの該音声特徴か ら口形特徴への変換を行う変換部と、種々の口形 特徴と具体的な口形状を表現する口形状パラメー タとを対応付けた変換テーブルと、前記変換部で 得られる各音業ごとの口形特徴に対応する口形状 パラメータを前記変換テープルから取出す口形状 パラメータ取得部と、一定時間間隔の画像系列と して与えられる動画像を生成するために該口形状 パラメータ取得部から得られる口形状パラメータ の値の出力を前記音声合成部から与えられる各音 素ごとの持続時間に従って制御するための時間調 整部と、核時間調整部の制御のもとに前記口形状 パラメータ取得部から出力される口形状パラメー タの値に従って画像を生成するための画像生成部 とをに加えて、前記時間調整部の出力に従ってあ る音楽から次の音楽への遷移を検出するための遷 移検出部と、前記画像生成部で用いられる口形状 パラメータの値を少なくとも1フレーム時間以上 保持することが可能なメモリと、該メモリに保持 されている口形状パラメーダの値と前記口形状パ ラメータ取得部より与えられる口形状パラメータ の値との中間値を求める口形状パラメータ修正部 とを更に隨え、ある音楽から次の音楽への遷移時 に中間的な口形状を生成して滑らかな口形状変化 を有する顚動画像を生成することにある。

(実施例1)

第1図は、本発明における第1の実施例を説明するためのプロック図である。入力情報としては、キーボード或いは磁気ディスク等のファイル装置から得られる文字列(文章)を考える。第1図において、1は音声合成部、2は時間を部、3は音声特徴から口形特徴への変換テーブル、5は口形状パラメータ取得部、6は画像生成部、10はゲート、900は文字列入力用の端子、901は画像出力用の端子である。

次に各部の動作について説明する。音声合成部 1 は入力された文字列に対応した音声出力を合成 する部分である。音声合成に関しては従来各種の 方式が提案されているが、ここでは、口形状生成 との整合性が優れているという点から、声道モデ ルとしてKlatt 型ホルマント音声合成器を用いた 既存の音声規則合成手法の利用を想定している。 この手法に関しては、(山本城一、機口宜男、流

次に時間調整部2は、音声合成部1より得られる各音楽毎の持続時間(第1番目の音素の持続時間をtaとする)に基づいて、画像生成部6への口形状パラメータの受渡しを制御するためのものである。すなわち、テレビジョン信号として画像(特に動画像)を出力するためには、例えばNTSC方

式の場合年秒30フレーム(1フレーム当り1/30秒) であり、1/30秒毎の情報に直して西像を生成する 必要がある。時間調整部2の詳しい動作について は後述する。

次に、音朗特徴から口形特徴への変換部3では、音声合成部1から得られる音韻特徴に基づいて、 該当音素に対応する口形特徴への変換を行う。口 形特徴としては、例えば、(1)口の開き具合(かなり開いている~完全に閉じている)、(2)唇の丸め 具合(丸めている~機に引いている)、(3)下顎の高 さ(上がっている~下がっている)、(4)舌の見え具 合、を考える。各種の音楽に対して、人間が実際 にどう発声しているかに関する観察に基づいて、 音韻特徴と口形特徴との対応を規則化している。

例えば、 "konnichiwa" という文章が入力された場合、

(無度) lv0 lh4 jaw0

k lv2 lhx jaw2 tbck

o 1v2 1h1 jaw2 -

:

のような形で口形特徴への変換がなされる。ここで、Iv,1h,Jaw は各々口の開き具合、唇の丸め具合、下顎の高さを示しており、数字は程度を表している。x は程度が前後の音楽によって決められることを示している(この場合、舌が奥の方にわずかに見えることを表している)。

ここで、時間調整部2の動作について詳しく説明する。第3図は時間調整部2の動作を説明するためのプロック図である。第3図において、21は遅延部、22は大小判定部、23、24はメモリ、25、26は加算器、27はスイッチ、28、29は分岐、30は時間正規化部、201、202は大小判定部22の出力

做の各々について、実際に人が発声をする時の口形状に対する計測結果を参考にして前もって定められた上記パラメータ P . ~ P . ・ Q . ~ Q . ・ h . , h . ・ θ . ・ θ . の値の組をテーブルの形で保持しておく。

口形状パラメータ取得部5では、音声特徴から口形特徴への変換部3より得られる該当音素に対する口形特徴に対して、口形特徴から口形状パラメータへの変換テーブル4を参照して、該当音素に対する口形状パラメータの値の組を取得する。

ゲート10は、該当音素に対する上記口形状パラメータを画像生成部6に送るか否かを制御するためのものであり、時間調整部2から指示された回数(この回数に1/30秒を乗じた値が、該当音素に対する口形状の表示時間となる)だけ、上記口形状パラメータを画像生成部6に送る。

画像生成部 6 はゲート10を介して口形状パラメータ取得部 5 より送られてくる1/30秒毎の口形状パラメータに基づいて口形状画像の生成を行う。 必要に応じて餌全体を含めた画像の生成を行う。

線、902 は初期リセット用の端子、 903は定数(1 /30)入力用端子、920、921はスイッチ27に関わる 嫡子である。次に各部の動作について説明する。 メモリ23は、1番目の音楽までの合計の持続時間 Σι: を蓄えておくためのメモリである。面像合 成を始める前に、端子 902より与えられる初期り セット信号で客がセットされる。音声合成部1か ら1番目の音素の持続時間が与えらると、加算器 25により、メモリ23に蓄えられた!一1番目の音 素までの合計の持統時間Σ t , と t , との和Σ t_i が求められる遅延部21は、「-1番目の音素まで の合計の持続時間∑ t , を、【+】番目の音素に 対する処理に入るまで蓄積する働きをする。時間 正規化部30では、遅延部21の出力Σί;に対し、 1/30×N≤∑ t; <1/30×(N+1) を満足するNを 求め、1/30×Nの値を出力する。. ここで、Nは整 数、また、1/30は1フレームの時間1/30秒を与え る定数である。スイッチ27は、1番目の音楽に対 する処理に入る時に、大小判定部22からの出力線 202により端子 920の側に接続される。この時、

加算器26により、時間正規化部30の出力1/30×N と定数1/30との和tが計算される。大小判定部22 では、このtの値と、tiの値との大小を比較し、 τ≦Σιιの場合には、出力線 201に、またι> Σι, の場合には出力線 202に信号を出力する。 t>Σtiの場合は、「番目の音素の持続時間が 終了したことを意味し、出力線 202を介して、音 声合成部1へ1+1番目の音素に関する情報を出 力するための指示、メモリ24へ内容をリセットす るための指示、スイッチ27へ端子 920へ接続する ための指示、遅延部21へ遅延されていたΣιιの 値を出力するための指示がなされる。メモリ24は、 加算器26の出力を一時的に蓄えておくためのもの である。スイッチ27は t ≦ ∑ t; が成立つ間端子 921に接続されており、加算器26により、順次、 今までのしに1/30を加えたものを新たなしにする 操作が行われる。以上により、 t ≤ Σ t ; が成立 つ間、大小判定部22より出力線 201に信号が出力 され、この信号により第1図におけるゲート10が 制御されることにより、【番目の音素の持統時間

の間、「番目の音楽に対応する口形状パラメータ が西像生成部6に供給される。

以上が本発明の第1の実施例に対する説明である。ここで、第1の実施例の場合、「番目の音楽から「十1番目の音素に移る場合、「番目の音楽に対する口形状パラメータから、「十1番目の音楽に対することになる。両者のこれでも、「一名で、一名で、一名で、「一名で、「一名」で、「一、「一、「一、「一、「一、「一、

(実施例2)

第4図はこの要求を満足するための本発明の第2の実施例を説明するためのブロック図である。 第4図において、7は口形状パラメータ修正部、 8は選移検出部、9はメモリ、40はスイッチ、 910、911はスイッチ40に関わる端子、他は第1図 に同様である。次に新たに加わった部分の動作を

説明する。

遷移検出部8は、ある音素(例えば【番目の音 素)から次の音楽(『+1番目の音楽)への運移 を検出するためのものである。第5回は本発明に よる遷移検出部8の動作を説明するためのプロッ ク図であり、81はカウンタ、82は判定回路、210. 211 は出力線である。カウンタ81は、大小判定部 22からの出力線 202に信号が出力された時に0に リセットされる。また、大小判定部22において出 力線201 に信号が出力されるごとに1ずつカウン トアップする。 判定回路82では、カウンタ81の出 力が"1"であるか否かを判定し、"1"の時に は、ある音素から次の音素への湿移が生じたとい うことであるので、出力線210 に信号を出力する。 ー方、 2 以上の時には、現在の音楽が持続してい るということであるので、出力線211 に信号を出 カする.

メモリ9は、前フレームの西像を合成するため に用いられた口形状パラメータを少なくとも1フ レーム期間若えておくためのメモリである。口形 状パラメータ修正部では、メモリ9に置えられて いた前フレームにおける口形状パラメータと、口 形状パラメータ取得部5より与えられる現在の音 素に対する口形状パラメータとに基づいて、例え ば両者の中間値を求めて、現フレームの画像を合 成するための口形状パラメータとする働きをする。 スイッチ40は、遷移検出部から出力線210, 211の いずれに信号が出力されるかによって、端子910 或いは911 に接続され、端子 910に接続された時 には、口形状パラメータ修正部7より得られる2 つの音器に対する口形状パラメータの中間値を、 また、端子911 に接続された時には現在の音楽に 対する口形状パラメータを、画像生成部6に渡す, 以上の例では、ある音素の口形状パラメータと次 の音素の口形状パラメータとの中間値は1フレー ム分しか生成されないが、例えばカンウタ82の値 に応じて何段階かの中間値を生成することにより、 より滑らかな口形状変化を実現することも可能で

以上述べたように、本発明は文字列として表現

特別平2-234285(プ)

される文章を入力した場合にこれに対応した口形となっても関いている。一番では、力力を開発している。一番では、力力を開発している。一番では、各音をは、力力を開発している。一番では、各音を関いている。一番では、本発明における音を関える。といる。一番では、大力音を開発した。一番である。

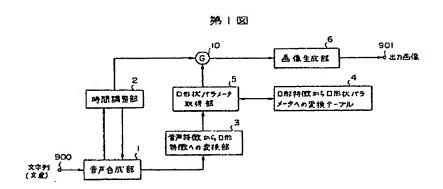
(発明の効果)

以上のように、本発明により、文字列として表現される文章を入力として音声出力との的確な対応付けがなされ、かつ、各音素の持続時間に合わせた口形状変化を有する、従って音声出力とのマッチングのとれた自然な口形状変化を有する動画像を合成することが可能である。

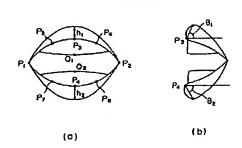
文章入力に対して、今まで音声を合成するのに とどまっていたのに対し、本発明では、音声との マッチングのとれた自然な口形状変化を有する動 画像まで容易に出力できるようになる。従って、本発明は実写を必要とせずにリアルな動画像を生成する用途(例えば、放送番組や映画の製作)、音声及び画像による自動応答装置、マン・マシーン・インタフェースの手段としての利用、文章から音声及び動画像へのメディア変換等に適用可能であり、その効果が極めて大である。

4. 図面の簡単な説明

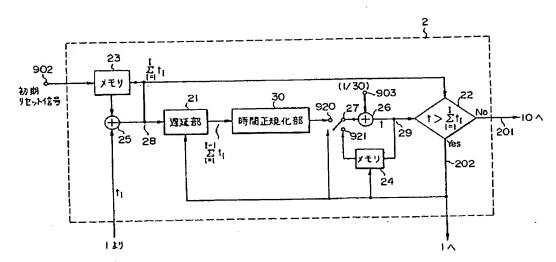
第1図は本発明の第1の実施例に対応するブロック図、第2図は口形状を表現するためのパラメータの例を示した図、第3図は本発明における時間調整部2の動作の一例に対応するブロック図、第4図は本発明の第2の実施例に対応するブロック図、第5図は本発明の第2の実施例における選移験出部8の動作の一例に対応するプロック図である。

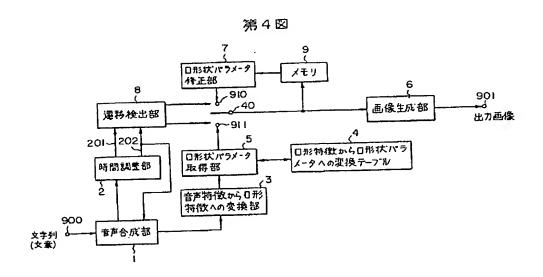


第2図

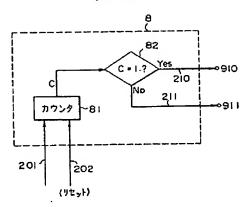


第3図

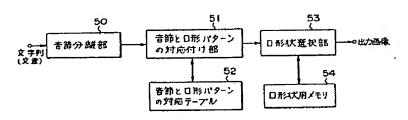




第5図



第6図



第1頁の続き ②発 明 者 樋 口 宜 男 東京都新宿区西新宿2丁目3番2号 国際電信電話株式会 社内

This Page is Inserted by IFW Indexing and Scanning Operations and is not part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

BLACK BORDERS

IMAGE CUT OFF AT TOP, BOTTOM OR SIDES

FADED TEXT OR DRAWING

BLURRED OR ILLEGIBLE TEXT OR DRAWING

SKEWED/SLANTED IMAGES

COLOR OR BLACK AND WHITE PHOTOGRAPHS

GRAY SCALE DOCUMENTS

LINES OR MARKS ON ORIGINAL DOCUMENT

REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY

IMAGES ARE BEST AVAILABLE COPY.

☐ OTHER: ___

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.